

Commonsense reasoning, commonsense knowledge, and the SP theory of intelligence

J Gerard Wolff*

September 27, 2016

Abstract

This paper describes how the *SP theory of intelligence*, outlined in an Appendix, may throw light on aspects of commonsense reasoning (CSR) and commonsense knowledge (CSK) (together shortened to CSRK), as discussed in another paper by Ernest Davis and Gary Marcus (DM). The SP system has the generality needed for CSRK: Turing equivalence; the generality of information compression as the foundation for the SP system both in the representation of knowledge and in concepts of prediction and probability; the versatility of the SP system in the representation of knowledge and in aspects of intelligence including forms of reasoning; and the potential of the system for the seamless integration of diverse forms of knowledge and diverse aspects of intelligence. Several examples discussed by DM, and how they may be processed in the SP system, are discussed. Also discussed are current successes in CSR (taxonomic reasoning, temporal reasoning, action and change, and qualitative reasoning), how the SP system may promote seamless integration across these areas, and how insights gained from the SP programme of research may yield some potentially useful new ways of approaching these topics. The paper considers how the SP system may help overcome several challenges in the automation of CSR described by DM, and how it meets several of the objectives for research in CSRK that they have described.

Keywords:

*Dr Gerry Wolff, BA (Cantab), PhD (Wales), CEng, MBCS, MIEEE; CognitionResearch.org, Menai Bridge, UK; jgw@cognitionresearch.org; +44 (0) 1248 712962; +44 (0) 7746 290775; *Skype*: gerry.wolff; *Web*: www.cognitionresearch.org.

1 Introduction

“AI has seen great advances of many kinds recently, but there is one critical area where progress has been extremely slow: ordinary commonsense.” So say Ernest Davis and Gary Marcus [5, p. 92], illustrating the point with questions that are easy for people to answer but can be difficult for artificial systems, such as “Who is taller, Prince William or his baby son Prince George” and “Can you make a salad out of a polyester shirt?” (*ibid.*).

Questions like those can be challenging for AI but the *SP theory of intelligence* and its realisation in the *SP computer model*, outlined in Appendix A, may provide a way forward. This paper describes how the SP theory may throw light on aspects of commonsense reasoning and commonsense knowledge, hereinafter abbreviated as ‘CSR’ and ‘CSK’, respectively, and, together, as ‘CSRK’. The discussion focusses mainly on problems in AI that may be seen to fall broadly within the area of CSRK, without attempting to give a precise definition of that concept.

The following sections (2 to 5) describe aspects of the SP system that relate to CSRK. These ideas feed in to later sections (from Section 7) which consider aspects of CSRK and how they may be modelled in the SP system, drawing mainly on examples and discussion in [5].

To avoid unnecessary repetition of information, aspects of the SP system that have been described in other publications will be outlined in this paper, with examples reduced to their essentials, and with references to where further information may be found.

As indicated in Appendix A, a central part of the SP programme of research has been the drive to discover or invent a conceptual framework that would simplify and integrate observations and concepts across artificial intelligence, mainstream computing, mathematics, and human perception and cognition. This is in accord with Occam’s Razor, widely accepted as a key principle in good science.

In the quest for simplification and integration, two closely-related principles have emerged:

- That, as a working hypothesis, much of computing and cognition may be understood as information compression via the matching and unification of patterns (ICMUP);
- More specifically, that much of computing and cognition may be understood as information compression via *multiple alignment* (ICMA), where “multiple alignment” is a concept borrowed and adapted from bioinformatics. The versatility of this new version of multiple alignment concept, demonstrated in earlier publications and outlined below, suggests that it has the potential to be the “double helix” of intelligence—as significant for an understanding

of intelligence, broadly construed, as is DNA for biological sciences.¹

2 Why the SP system may provide a foundation for CSRK

This and the following sections provide some reasons for thinking that the SP system may provide a foundation for CSRK. In summary:

- *Turing equivalence.* The SP system is, probably, Turing-equivalent in the sense that it can in principle perform any computation within the scope of a universal Turing machine ([14, Chapter 4], [19, Section 6.6]) but it provides much of the human-like intelligence that, as Turing recognised [10, 11], is missing from the universal Turing machine. Thus it has the kind of generality needed for CSRK, and its strengths in AI give it a head start in modelling the intelligence of CSRK.
- *Information compression via multiple alignment.* As noted in the Introduction and in Appendix A, a central part of the SP system is ICMUP and, more specifically, ICMA. An important point here is that the remarks that follow apply to ICMUP and ICMA, not information compression in general. The latter encompasses many techniques that do not exhibit the strengths and potential of the SP system in reasoning and other aspects of AI.

For CSRK, ICMUP and ICMA have a two-fold significance:

- *Generality in the representation of knowledge.* The generality of ICMUP and ICMA suggests that, in principle, *any* kind of knowledge may be represented in the SP system. Of course, “any kind of knowledge” may include forms of knowledge that people would judge to be poor representations of “reality” as we perceive it.

But, as noted in Appendix A, unsupervised learning in the SP computer model as it has been developed to date, conforms to the DONSVIC principle, meaning that it creates knowledge structures that people regard as natural and which yield relatively high levels of information compression. It appears that such forms of knowledge are those that are most relevant to CSRK.

¹Unless otherwise stated, the expression “multiple alignment” in this paper will mean multiple alignment as it has been developed in the SP programme of research, not multiple alignment as that term is understood in bioinformatics.

- *Generality in prediction and probability.* As noted in Appendix A, the intimate connection that is known to exist between information compression and concepts of prediction and probability [6] means that inference and probabilities are central in the workings of the SP system, in keeping with the probabilistic nature of CSR.

The generality of these principles and the probabilistic nature of much CSR lends support to the SP system a possible foundation for CSRK,

- *Versatility of the SP system.* As outlined below, the SP system demonstrates versatility in the representation of knowledge (Section 3), versatility in aspects of intelligence (Section 4), and, more specifically, versatility in forms of reasoning (Section 5). This versatility of the SP system suggests that it may prove useful in modelling CSRK.
- *Seamless integration of forms of knowledge and aspects of intelligence.* The use of one simple format for representing several forms of knowledge and one relative simple framework—multiple alignment—for several aspects of intelligence, is likely to facilitate the seamless integration of diverse kinds of knowledge and diverse aspects of intelligence. Such integration appears to be essential for realistic modelling of CSRK.

It is pertinent to note that similar principles argue for the possibility that the SP system may be developed into a *universal framework for the representation and processing of diverse kinds of knowledge* (U FK) [18, Section III-B].

3 Versatility in the representation of knowledge

The quest for simplification and integration of observations and concepts in AI and related areas has led to the creation of a system that combines relative simplicity in its organisation with versatility in the representation of diverse forms of knowledge and versatility in diverse aspects of intelligence, as outlined in the sections that follow.

As noted in Appendix A, it is envisaged that all kinds of knowledge in the SP system are to be represented with SP *patterns* meaning arrays of atomic symbols in one or two dimensions. Despite their simplicity, SP patterns, within the multiple alignment framework, have proved to be effective in representing several forms of knowledge, any of which may serve in CSK. These include: the syntax of natural language; class hierarchies, class heterarchies (meaning class hierarchies with cross-classification); part-whole hierarchies; discrimination networks and trees; entity-relationship structures; relational knowledge; rules for use in reasoning; patterns

in one or two dimensions; images; structures in three dimensions; and procedural knowledge. There is more detail throughout [14] and [15], and there are references to further sources of information in [18, Section III-B]. Some examples are shown below.

4 Versatility in aspects of intelligence

Despite the essential simplicity of the SP system, it demonstrates strengths and potential in several aspects of intelligence including: unsupervised learning; natural language processing; fuzzy pattern recognition; recognition at multiple levels of abstraction; best-match and semantic forms of information retrieval; several kinds of reasoning (more in Section 5, below); planning; and problem solving ([14, Chapters 5 to 9], [15, Section 10]).

4.1 Pattern recognition with class-inclusion relations, part-whole relations, and inheritance of attributes

To illustrate aspects of intelligence and knowledge representation in the SP system, Figure 1² shows how, via the building of a multiple alignment, the SP computer system may model the recognition of an unknown plant at multiple levels of abstraction and in terms of the parts and sub-parts of flowering plants. It also illustrates a very useful and widely-used form of inference that appears to be a prominent feature of CSR (more below).

For the creation of this multiple alignment, and others, the SP computer model was supplied with a set of New patterns shown in column 0 of the figure and a relatively large set of Old patterns including those shown in columns 1 to 6. The New patterns—the one-symbol pattern ‘`has_chlorophyll`’ and the multi-symbol patterns ‘`<stem> hairy </stem>`’, ‘`<petals> yellow </petals>`’, ‘`<stamens> numerous </stamens>`’, and ‘`<habitat> meadows </habitat>`’—describe the features of some unknown plant.³ The Old patterns describe the structures and attributes of different classes of plant. As with the order of Old patterns across the rows in each of the two multiple alignments in Figure 7, the order of the Old patterns across columns 1 to 6 in Figure 1 is quite arbitrary and with no significance.

²Compared with the two multiple alignments shown in Figure 7, this multiple alignment has been rotated by 90°, with the New pattern in column 0 and Old patterns in columns 1 to 6. The choice between these two ways of displaying multiple alignments depends purely on what fits best on the page.

³These New patterns may be supplied to the SP computer model in any order, not only the order shown in column 0 of the figure.

0	1	2	3	4	5	6
	<species>					
	acris					
	<genus>					<genus>
	Ranunculus					Ranunculus
					<family>	<family>
					Ranunculaceae	Ranunculaceae
				<order>	<order>	
				Ranunculales	Ranunculales	
			<class>	<class>		
			Angiospermae	Angiospermae		
		<phylum>	<phylum>			
		Plants	Plants			
		<feeding>				
has_chlorophyll	has_chlorophyll	has_chlorophyll				
	photosynthesises					
	<feeding>					
	<structure>	<structure>				
		<shoot>				
<stem>	<stem>	<stem>				
hairy	hairy					
</stem>	</stem>	</stem>				
	<leaves>	<leaves>				
	compound					
	palmately_cut					
	</leaves>	</leaves>				
		<flowers>			<flowers>	
					<arrangement>	
					regular	
					all_parts_free	
					</arrangement>	
	<sepals>	<sepals>			<sepals>	
	not_reflexed					
	</sepals>	</sepals>			</sepals>	
<petals>	<petals>	<petals>			<petals>	<petals>
					<number>	<number>
					five	five
					</number>	</number>
	<colour>	<colour>			<colour>	
yellow	yellow					
</petals>	</colour>	</colour>			</colour>	
	</petals>	</petals>			</petals>	</petals>
					<hermaphrodite>	
<stamens>	<stamens>	<stamens>			<stamens>	
numerous	numerous	numerous			numerous	
</stamens>	</stamens>	</stamens>			</stamens>	
		<pistil>			<pistil>	
		ovary			ovary	
		style			style	
		stigma			stigma	
		</pistil>			</pistil>	
		</hermaphrodite>			</hermaphrodite>	
		</flowers>			</flowers>	
		</shoot>			</shoot>	
		<root>			<root>	
		</root>			</root>	
		</structure>			</structure>	
<habitat>	<habitat>	<habitat>				
meadows	meadows					
</habitat>	</habitat>	</habitat>				
	<common_name>	<common_name>				
	Meadow					
	Buttercup					
	</common_name>	</common_name>				
	<food_value>	<food_value>			<food_value>	
					poisonous	
	</food_value>	</food_value>			</food_value>	
	</phylum>	</phylum>				
		</class>			</class>	
					</order>	
					</family>	</family>
	</genus>	</genus>				
	</species>					

Figure 1: The best multiple alignment created by the SP model, with a small set of New patterns (in column 0) that describe some features of an unknown plant, and a set of Old patterns, including those shown in columns 1 to 6, that describe different categories of plant, with their parts and sub-parts, and other attributes. From Figure 16 in [15], reproduced with permission.

The multiple alignment in the figure—the best of those created by the SP computer model—shows that the unknown plant is a Meadow Buttercup (species *acris* in column 1), in the genus *Ranunculus* (column 6), which is in the family Ranunculaceae (column 5), the order Ranunculales (column 4), and so on.

Identification is achieved via attributes at several different levels in the class hierarchy, including attributes which have a part-whole hierarchical structure such as: ‘flowers’ in column 3; broken down into ‘sepals’, ‘petals’, ‘stamens’, and other attributes, in column 5; with further details of petals given in column 6 (the number of petals which in this case is ‘five’), and in column 1 (the colour of the petals which in this case is ‘yellow’). The figure illustrates an important feature of the SP system: that there can be seamless integration of class-inclusion relations with part-whole relations, as discussed in Section 9.1.2 below.

An aspect of recognition via the SP system that is not illustrated in Figure 1 is that, like people, the system has a robust ability to recognise patterns despite errors of omission, commission, and substitution in the pattern or patterns that are to be recognised. Examples may be seen in [14, Section 6.2] and [15, Sections 4.2.2 and 5.3].

4.2 Inheritance of attributes

As already indicated, recognition via multiple alignment, with class-inclusion relations and with part-whole relations, provides a means of making a type of inference that is bread-and-butter in everyday reasoning and everyday thinking. This type of inference, which is called “inheritance of attributes” in object-oriented programming⁴ and relates to “Prediction by partial matching” in information compression,⁵ means predicting the unseen parts of a pattern that has been recognised, or a set of such patterns.

With recognition via multiple alignment, this can be done at any of the levels in the multiple alignment, except column 0. Thus, for example, in the multiple alignment in Figure 1, we may infer that the plant *Ranunculus acris* has sepals that are not reflexed and leaves that are compound and palmately cut (the ‘species’ level in column 1), that the plant nourishes itself via photosynthesis (the ‘phylum’ level in column 2), and that it is poisonous (the ‘family’ level in column 5). With more detailed information in the SP patterns, many more such inferences would be possible.

The intimate relation between pattern recognition and inference via inheritance of attributes illustrates a general truth about the SP system: that there is poten-

⁴See, for example, “Object-oriented programming”, *Wikipedia*, bit.ly/20Rx76M, retrieved 2016-08-08.

⁵See, for example, “Prediction by partial matching”, *Wikipedia*, bit.ly/1BUtAYo, retrieved 2016-08-08.

tial for the seamless integration of different aspects of intelligence, an integration that is a prominent feature of human intelligence and appears to be essential in any any system that aspires to achieve the versatility and adaptability of human intelligence. We shall return to this point in Section 6.

5 Versatility in reasoning

Although reasoning is an aspect of intelligence (Section 4), it has been given a section to itself because of the versatility of the SP system in this area and because it is a key part of CSR.

In reasoning, strengths and potential of the SP system include ([14, Chapter 7], [15, Section 10]): one-step ‘deductive’ reasoning; chains of reasoning; abductive reasoning; reasoning with probabilistic networks and trees; reasoning with ‘rules’; nonmonotonic reasoning and reasoning with default values; Bayesian reasoning with “explaining away” (as discussed by Judea Pearl in [9, Sections 1.2.2 and 2.2.4]); causal reasoning; and reasoning that is not supported by evidence. As we have seen in Section 4.2, the SP system also supports inference via inheritance of attributes. It appears that there is also potential for spatial reasoning [17, Section IV-F.1], and for what-if reasoning [17, Section IV-F.2].

It would not be either feasible or appropriate to reproduce everything in this area that has been published before. Instead, the next two subsections describe two aspects of reasoning as it may be developed in the SP system.

5.1 Nonmonotonic reasoning and reasoning with default values

A popular example of nonmonotonic reasoning is how, if we are told that ‘Tweety’ is a bird, we would normally infer, on the strength of the default assumption that most birds can fly, that it is likely that Tweety can fly. But if, later, we are told that Tweety is a penguin, we will revise our thinking and conclude with some confidence that Tweety cannot fly. This conflicts with classical logic, where later evidence should not alter an earlier conclusion—but it is entirely consistent with everyday thinking and CSR.

With an appropriate store of Old patterns, and the New pattern ‘bird Tweety’ (which may be interpreted as “Tweety is a bird”), the best of the multiple alignments created by the SP computer model is the one shown in Figure 2 ([15, Section 10.1], [14, Section 7.7]). This has a relative probability of 0.66.

In addition to the multiple alignment in Figure 2, the SP computer model creates two other multiple alignments that provide an interpretation for all the symbols in the New pattern. The second-best multiple alignment is similar to the

0	1	2	3
			Default
		Bd -----	Bd
bird -----		bird	
	name ---	name	
Tweety -	Tweety		
	#name --	#name	
		f -----	f
			canfly
		#f -----	#f
		warm-blooded	
		wings	
		feathers	
		...	
		#Bd -----	#Bd
			#Default
0	1	2	3

(a)

Figure 2: The best multiple alignment formed by the SP computer model with ‘bird Tweety’ in New and patterns in Old as described in the text. The relative probability of this multiple alignment, compared with alternatives formed at the same time, is 0.66.

first one except that the pattern in column 3 is replaced by the pattern ‘0 ostrich Bd f cannotfly #f #Bd ... #0’, while the third-best multiple has the the pattern ‘P penguin Bd f cannotfly #f #Bd ... #P’ in that position. These two multiple alignments have relative probabilities of 0.22 and 0.12, respectively.

From these three multiple alignments we may conclude that, in accordance with commonsense, it is most likely that Tweety is a bird that can fly but that it is possible but less likely that Tweety is an ostrich or a penguin, and that in both of those cases, Tweety would not be able to fly.

Now, if we run the SP computer model again but with ‘penguin Tweety’ as the New pattern (which may be interpreted as “Tweety is a penguin”), and with the same Old patterns as before, the system produces only one multiple alignment that provides an interpretation for all the symbols in the New pattern, the multiple alignment shown in Figure 3. Since there are no rivals for this multiple alignment, its relative probability is calculated as 1.0. From this, we may make the inference that, as a penguin, Tweety certainly cannot fly.

0	1	2	3
			P
penguin	-----		penguin
		Bd	----- Bd
		bird	
	name ---	name	
Tweety	-- Tweety		
	#name --	#name	
		f	----- f
			cannotfly
		#f	----- #f
		warm-blooded	
		wings	
		feathers	
		...	
		#Bd	----- #Bd
			...
			#P
0	1	2	3

Figure 3: The best multiple alignment formed by the SP computer model with the New pattern ‘penguin Tweety’ and the same Old patterns as before. The relative probability of this multiple alignment is 1.0.

5.2 Spatial reasoning

As mentioned in Appendix A, it is envisaged that knowledge may be represented in the SP system with two-dimensional patterns as well as 1D patterns, and that the SP computer model will be generalised to work with 2D patterns. Amongst other things, this should facilitate the learning of 3D structures and the application of such structures in spatial reasoning.

With regard to the learning of 3D structures, the SP system may create a 3D model of an object or other structure by stitching together partially-overlapping pictures of the structure, taken from different angles, as illustrated schematically in Figure 4. The process of recognising partial matches between pictures would be done via a search for good full and partial matches between patterns, a process which is central in the workings of the SP system [16, Sections 6.1 and 6.2]. Creating a 3D digital model of an object from overlapping pictures is already done by commercially-available systems. And, in a similar way, Google Streetview uses overlapping pictures to build what are, in effect, 3D digital models of street plans.

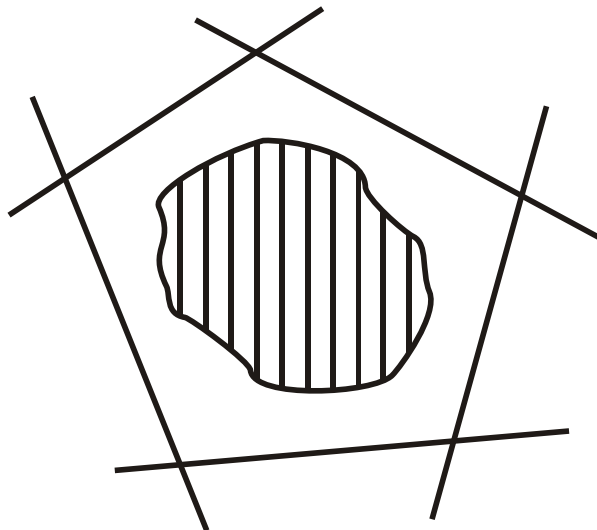


Figure 4: Plan view of a 3D object, with each of the five lines around it representing a view of the object, as seen from the side.

Development of the SP system along the lines just outlined should open up possibilities for spatial reasoning, as outlined in [17, Section IV-F.1]. For example, it may prove possible to do such things as planning how furniture may be arranged in a room by digital ‘manipulation’ of digital models of the furniture and the room, in much the same way that people sometimes plan the furnishing of a room by

trying out various arrangements of wooden or card-board representations of the furniture and the room. In a similar way, it may prove possible to discover whether or not a large or awkwardly-shaped object may go through a given space by trying things out digitally, instead of with the real thing.

6 Seamless integration of diverse forms of knowledge and diverse aspects of intelligence

Three features of the SP system suggests that it should facilitate seamless integration of diverse kinds of knowledge and seamless integration of diverse aspects of intelligence (including several forms of reasoning):

- The adoption of one simple format for all kinds of knowledge;
- That one relatively simple framework—multiple alignment—is central in all kinds of processing;
- That the relatively simple format for knowledge and the multiple alignment framework for processing knowledge provides for the representation of several different kinds of knowledge (Section 3) and several aspects of intelligence (Section 4, with several forms of reasoning (Section 5).

In a relatively simple form, those two kinds of integration may be seen in the example discussed in Section 4.1, Figure 1:

- Class-inclusion relations and part-whole relations work together in combination in the representation of knowledge without awkward incompatibilities.
- Pattern recognition (an aspect of general intelligence) and inheritance of attributes (a form of reasoning) are intimately related in what is, in effect, one type of operation.⁶

For the understanding of natural language and the production of language from meanings, it is likely to be helpful if there is seamless integration of syntax and semantics, and it seems likely that this will be facilitated by representing both of

⁶What Conan Doyle, via his characters, calls “deduction” may, very often, equally well be seen as pattern recognition, witness Sherlock Holmes’ remarks to Dr Watson: “As to your practice, if a gentleman walks into my rooms smelling of iodoform, with a black mark of nitrate of silver upon his right forefinger, and a bulge on the right side of his top-hat to show where he has secreted his stethoscope, I must be dull, indeed, if I do not pronounce him to be an active member of the medical profession.” From “A Scandal in Bohemia”, *The Complete Sherlock Holmes*, Starbooks Classics Publishing, Kindle Edition.

them with SP patterns, and by processing both of them together via the building and manipulation of multiple alignments. Some preliminary examples from the SP computer model show how this kind of integration may be achieved with both the understanding and production of natural language [14, Section 5.7]. There is clear potential with the SP system for the comprehensive integration of the syntax and semantics of natural language.

Seamless integration of diverse kinds of knowledge and diverse aspects of intelligence and diverse forms of reasoning is important for the acceptance of the SP theory as a theory of CSRK since, as a matter of ordinary experience, CSR, or ‘everyday’ reasoning, means a willingness to use any and all relevant forms of knowledge, and a willingness to be flexible in one’s thinking—to use any and all forms of reasoning with other aspects of intelligence where appropriate. More generally, it appears that this kind of integration is essential for the achievement of human-like intelligence, with all its versatility and adaptability.

7 Father and son, and other examples

This and the following main sections discusses aspects of CSRK in the light of what has been said about the SP system earlier in this paper and in Appendix A. The discussion focuses mainly on what Davis and Marcus (DM) have said about CSRK in [5], and uses several of their headings.

In their introduction to CSRK, DM give some examples of natural language and everyday situations that are easy for people to understand but can be difficult for AI systems:

- “Who is taller, Prince William or his baby son Prince George?” DM say “... if you see a six-foot-tall person holding a two-foot-tall person in his arms, and you are told they are father and son, you do not have to ask which is which.” (p. 92).
- “Can you make a salad out of a polyester shirt?” DM say “If you need to make a salad for dinner and are out of lettuce, you do not waste time considering improvising by taking a shirt [out] of the closet and cutting it up.” (pp. 92–93).
- DM say “If you read the text, ‘I stuck a pin in a carrot; when I pulled the pin out, it had a hole,’ you need not consider the possibility [that] ‘it’ refers to the pin.” (p. 93).

Although preliminary work, mentioned in Section 6, shows how, in the SP system, syntactic and semantic knowledge can work together in the understanding

of natural language, more work would be needed to demonstrate an understanding of example sentences like those just shown. But the kinds of inferences needed for the understanding of those sentences are well within the scope of the SP system. It appears that all three of them depend largely on inheritance of attributes, discussed in Section 4.1:

- With the father and son example, height is the kind of attribute that would normally be associated with people of all kinds and with subclasses like ‘father’ and ‘son’. And that knowledge would suggest immediately, via inheritance of attributes, that the father would be taller than the son, especially since the son is described as a baby. Although the inference is probabilistic, the information that Prince George is a baby, the knowledge that most people have about Prince William, and the difficulty that a small person would have in holding a big person in their arms, would, very likely, rule out the possibility that the father is a dwarf and that the son is fully grown.
- The salad example depends more directly on inheritance of attributes: anything that goes into a salad must have the attribute ‘edible’, at any level in the hierarchy or hierarchies of classes in which it belongs. A polyester shirt clearly fails that test.
- In a similar way, the carrot and pin example depends, at least in part, on characteristics of carrots (they are relatively soft) and pins (they are normally made of something that is relatively hard, normally metal, and they are normally designed to stick into things) that may be inherited from any of the levels in the classes in which they belong. Also relevant is the meanings of the words in the phrase “stuck X in Y”, which implies that X would make a hole in Y.

8 Commonsense in intelligent tasks

This section considers a selection of topics considered by DM under the above heading.

8.1 The city councilmen refused the demonstrators a permit

As noted by DM (p. 93), the subtlety of natural language may be seen in a pair of example sentences presented by Terry Winograd in [12, p. 33]:

- *The city councilmen refused the demonstrators a permit because they feared violence.*

- *The city councilmen refused the demonstrators a permit because they advocated revolution.*

People naturally assume that, in the first sentence, “they” means the city councilmen, while in the second sentence, “they” means the demonstrators, but those two inferences may be problematic for AI systems.

As with the examples in Section 7, a full interpretation of sentences like these is beyond the scope of the SP computer model as it is now. But, as before, the key to the interpretation of these sentences, appears to be inheritance of attributes as discussed in Sections 7 and 4.1. It is generally known that, as a class, city councilmen will normally wish to maintain peace in their city, while a subset of the class “demonstrators” do sometimes advocate revolution. Inferences from that knowledge (via inheritance of attributes), with a knowledge of the meanings of the words “feared” and “advocated”, appears to be sufficient, in each of the two cases above, to disambiguate the pronoun “they”.

8.2 Computer vision

With respect to a photograph of a kitchen shown in [5, Figure 1], DM say:

“Many of the objects that are small or partially seen, such as the metal bowls in the shelf on the left, the cold water knob for the faucet, the round metal knobs on the cabinets, the dishwasher, and the chairs at the table seen from the side, are only recognizable in context; the isolated image would be difficult to identify.” (p. 94).

They go on to say that “The viewer infers the existence of objects that are not in the image at all.” (*ibid.*) and that “The viewer also infers how the objects can be used (sometimes called their ‘affordances’);” (*ibid.*).

As described in the next two subsections, these capabilities appear to be well within the scope of the SP system.

8.2.1 The importance of context in recognition

Although the parsing of natural language is not the same as visual recognition and scene analysis, it appears that similar principles apply in both cases [16, Section 4]:

- Consider, first, that in much the same way that the SP computer model can find good alternative parsings of an ambiguous sentence (illustrated in Figure 7), it can discover the two most plausible analyses of the ambiguous

phoneme sequence ‘ae i s k r ee m’ (which can be read as “ice cream” or “I scream”) [14, Section 5.2.1].

- Then, second, and more directly relevant to the present discussion, the provision of disambiguating context (in the phoneme equivalents of “I scream loudly” and “Ice cream is cold”) will tilt the preferred analyses in one direction or the other, in accordance with human intuitions ([16, Sections 7.4 and 7.5], [14, Section 5.2.2]).

8.2.2 ‘Seeing’ things that are not objectively present in an image

The SP computer model provides an account of some types of situation where people ‘see’ things that are not objectively present in an image:

- In much the same way that, in the example of recognition discussed in Section 4.1, we may infer that the unknown plant has features that were not in the New information supplied to the SP computer model (sepals that are not reflexed, leaves that are compound and palmately cut, that the plant nourishes itself via photosynthesis, and that it is poisonous), we may infer objects that are not in an image, and uses for an object that are not visible either.
- David Marr [7] describes two examples where people ‘see’ things that are not objectively present in an image: the sides of a triangle that can be seen in Kanizsa’s triangle although it is mostly empty space (*ibid.*, Figure 2-6); and the line in a photograph of a plant, where one leaf overlaps another leaf, that we can ‘see’ although there is nothing objective to mark it (*ibid.*, Figure 4-1 (a)).

As described in [16, Section 7.1], these things may be seen to be the result of a process of visual parsing that has effect of introducing boundaries between segments, although those boundaries are not objectively present in what is being parsed. Examples of such parsing of natural language with the SP computer model may be seen in [15, Figure 6 (a)] and [14, Chapter 5]. An example from the SP computer model showing the parsing of a one-dimensional analogue of Kanizsa’s triangle is shown in [20, Section XI-A.2].

8.3 Robotic manipulation

Although the SP concepts appear to be highly relevant to the development of intelligence in robots, as described in [17], no attempt has yet been made to address problems that DM describe like this:

“If a cat runs in front of a house-cleaning robot, the robot should neither run it over nor sweep it up nor put it away on a shelf. These things seem obvious, but ensuring a robot avoids mistakes of this kind is very challenging.” [5, p. 94].

That said, it seems likely that the SP system’s strengths and potential in diverse forms of reasoning (Section 5, [17, Section IV-F]) will help it to avoid the kinds of mistake sketched by DM.

9 Successes in automated commonsense reasoning

This section considers four areas where, as described by DM, there have been successes with other approaches, and compares them with what may be done with the SP system.

9.1 Taxonomic reasoning

As we saw in Section 4.1, taxonomic relations may be expressed with SP patterns, with pattern recognition at multiple levels of abstraction.

9.1.1 Basic relations

With regard to the three basic taxonomic relations described by DM:

- *An individual is an instance of a category.* This can be seen in the way that the unknown plant in the example shown in Figure 1 may be recognised as an instance of the species Meadow Buttercup, a member of the genus *Ranunculus*, a member of the family Ranunculaceae, and so on.
- *One category is a subset of another.* In the SP system, this kind of relationship is expressed using pairs of symbols like ‘<genus> ... </genus>’ which, in the figure, provide the connection between the pattern describing the concept ‘genus’ (column 6 in the figure) and the pattern describing the concept ‘species’ (column 1).
- *Two categories are disjoint.* In the SP system, this kind of relationship may be implicit in a collection of SP patterns. There appears to be no need for it to be marked explicitly.

9.1.2 Categories and properties

DM write that “Categories can ... be tagged with properties. For instance, `Mammal` is tagged as `Furry` [in their Figure 2].” [5, p. 95]. In this connection, an important feature of the SP system is that there is no distinction between categories and properties (otherwise known as classes and attributes), and, in a similar way, there is no distinction between ‘classes’ and ‘objects’, a distinction which is prominent in object-oriented programming. In the SP system, all such concepts are modelled with ‘patterns’ and ‘symbols’, although concepts like ‘category’ and ‘property’ may be used informally where appropriate.

There are three main advantages in removing these distinctions:

- It facilitates the seamless integration of class-inclusion relations with part-whole relations, as illustrated in Figure 1. Achieving that integration was one of the original motivations for the development of the SP system.⁷
- If a class can be an object (which is a feature of some object-oriented systems), then there is a need for the category ‘metaclass’ (a class of a class). This in turn points to the need for such categories as ‘metametaclass’, ‘metametametaclass’, and so on, in a rather unhelpful recursive loop [14, Section 6.4.3.1].
- It facilitates the representation of cross-classification and other relatively complex kinds of taxonomy outlined in the third point in Section 9.1.3.

9.1.3 Forms of inference

With regard to the taxonomic forms of inference discussed by DM:

- *Transitivity.* “Since `Lassie` is an instance of `Dog` and `Dog` is a subset of `Mammal`, it follows that `Lassie` is an instance of `Mammal`.” [5, p. 95]. In the SP system, relationships like those may be implicit in a set of patterns describing different categories of animal. They would not be encoded explicitly.
- *Default inheritance.* “A variant of [inheritance] is default inheritance; a category can be marked with a characteristic but not universal property, and a subcategory or instance will inherit the property unless it is specifically canceled.” (*ibid.*). As with saw in Section 5.1, the SP system can model reasoning with default assumptions.

⁷When I was employed in the software industry and working on the development of a ‘support environment’ for software engineers, it became clear that there was a need to integrate classes or versions of software products with the parts and sub-parts of such products, and that this was difficult to do with existing technologies.

- *Other taxonomies.* “Other taxonomies are less straightforward. For instance, in a semantic network for categories of people, the individual `GalileoGalilei` is simultaneously a `Physicist`, an `Astronomer`, a `ProfessorOfMathematics`, a `WriterInItalian`, a `NativeOfPisa`, a `PersonChargedWithHeresy`, and so on. These overlap, and *it is not clear which of these are best viewed as taxonomic categories and which are better viewed as properties*. In taxonomizing more abstract categories, choosing and delimiting categories becomes more problematic; for instance, in constructing a taxonomy for a theory of narrative, the membership, relations, and definitions of categories like `Event`, `Action`, `Process`, `Development`, and `Incident` are uncertain.” (*ibid.*, emphasis added).

The points that DM make here relate to two features of the SP system:

- With SP patterns within the multiple alignment framework, it is as straightforward to model cross-classification or class hierarchies as it is to model ordinary class hierarchies [14, Section 6.4].
- That “... it is not clear which [of the attributes of Galileo] are best viewed as taxonomic categories and which are better viewed as properties” (emphasised in the quotation) lends support to the feature of the SP system (noted in Section 9.1.2) that it avoids formal distinctions between such things as ‘categories’ and ‘properties’

9.2 Temporal reasoning

DM write:

“Representing knowledge and automating reasoning about times, durations, and time intervals is a largely solved problem. For instance, if one knows that Mozart was born earlier and died younger than Beethoven, one can infer that Mozart died earlier than Beethoven. ... Integrating such reasoning with specific applications, such as natural language interpretation, has been ... problematic. ... many important temporal relations are not explicitly stated in texts, they are inferred; and the process of inference can be difficult.” [5, p. 96].

No attempt has yet been made to apply the SP system to temporal reasoning, so the remarks here are tentative.

As we saw in Section 5.2, the SP system has potential for the building of 3D digital models of objects and environments, and for spatial reasoning with such models. The suggestion here, partly motivated by what people seem to do in reasoning about temporal relations, is that such reasoning may be done in a

manner that is similar to or the same as spatial reasoning, via the computational manipulation of digital objects representing periods of time. This does not solve the problem of interpreting natural language descriptions of a temporal reasoning problem, but it may throw some light on how temporal reasoning may be done after the natural language description has been translated into an appropriate form.

There is indirect support for this idea via the thinking behind Cuisenaire rods: wooden or plastic sticks in different colours and different lengths which are used as an aid in the teaching of arithmetic concepts to children, by showing how addition, subtraction, multiplication and division, and other arithmetic operations, may be understood visually.⁸ It seems possible that commonsense reasoning about temporal relations may be done, mentally or digitally, in a similar way.

9.3 Action, change, and qualitative reasoning

DM describe two other areas of success in automated commonsense reasoning [5, pp. 96–97]:

- *Action and change.* Modelling inferential processes related to actions, events, and change, with possibly over-simplified assumptions such as “... one event occurs at a time, and the reasoner need only consider the state of the world at the beginning and the end of the event, ...”, and “Every change in the world is the result of an event.”
- *Qualitative reasoning.* Modelling forms of qualitative reasoning such as how the price of a product influences the number of items that are sold, and how an increase in the temperature of a gas in a closed container leads to an increase in pressure.

As with temporal reasoning, no attempt has yet been made to apply the SP system to these areas, so remarks about them are tentative—and they are reserved for the following subsection.

9.4 Discussion

As DM say, there has been some success in the four areas discussed above (Section 9). But there is one main shortcoming, outlined in the next subsection. And insights gained from the SP programme of research suggest some potentially useful new ways of approaching these topics, discussed in the next but one subsection. However, the successes that DM have described may also have some useful things to say about the SP system and how weaknesses in the system may be overcome.

⁸See “Cuisenaire rods”, *Wikipedia*, bit.ly/2bOEQ6N, retrieved 2016-09-05.

9.4.1 Compatibility and integration

What appears to be the main shortcoming of the four areas of success in CSRK described by DM is that they have apparently been developed quite independently of each other or other aspects of intelligence. This is the kind of fragmentation in AI that Pamela McCorduck criticised so lucidly: “The goals once articulated with debonair intellectual verve by AI pioneers appeared unreachable ... Subfields broke off—vision, robotics, natural language processing, machine learning, decision theory—to pursue singular goals in solitary splendor, without reference to other kinds of intelligent behaviour.” [8, p. 417]. And the fragmentation is at odds with the central aim of the SP research, to promote simplification and integration of observations and concepts across a broad canvass (Appendix A).

The penalty of developing these four areas independently of each other is that there is little or no integration amongst them, and they appear to have little or nothing to say to each other. Since they all deal with aspects of CSRK, it is disappointing that there is no overarching conceptual framework or theory. And it would have been useful to have some insight into how the four areas might integrate with other aspects of intelligence, especially learning.

An example that illustrates the potential benefits of integration is “... the problem of integrating action descriptions at different levels of abstraction.” mentioned in DM’s section about “Action and change” [5, p. 96]—although the examples given by DM seem to represent part-whole relations rather than class-inclusion relations. Either way, research on action and change would probably benefit from ideas about taxonomic reasoning and, in particular, how they may be realised in the SP system.

9.4.2 Unsupervised learning as a guide to CSRK

Given the current state of development of the SP system, as expressed in the SP computer model, a potentially useful way of developing ideas in taxonomic reasoning, temporal reasoning, action and change, and qualitative reasoning, and in other aspects of AI, is something like what follows:

1. Develop unsupervised learning in the SP system to solve residual problems as outlined in [15, Section 3.3].⁹
2. With appropriate sets of New patterns, examine the best *grammars* (sets of Old patterns) created by unsupervised learning in the computer model, and

⁹The main problem at present is that the SP computer model does not learn intermediate levels of abstraction—but it appears that this problem is soluble. It would also be useful to generalise the computer model to work with patterns in two dimensions.

evaluate their strengths and weaknesses in the representations of knowledge, and in terms of inferential processes in different areas of CSR.

3. In the light of insights gained in Stage 2, revise and refine the SP computer model.
4. Iterate Stages 2 and 3 until residual problems have been solved.

The main reason for adopting this approach is that can be difficult or impossible to gain the necessary insights in any other way—in much the same way that an aircraft engineer, for example, would find it difficult or impossible to understand thoroughly how a new kind of aircraft will behave, except via the creation of prototypes, with the development of computer models, validated and refined in the light of data from the prototypes.

10 Challenges in automating commonsense reasoning

In a section with the heading above [5, pp. 97–99], DM describe five challenges for the automation of commonsense reasoning, discussed in the following subsections.

10.1 Many domains are poorly understood

“... many of the domains involved in commonsense reasoning are only partially understood or virtually untouched.” [5, p. 97].

This is true and DM’s paper [5] does a valuable service in highlighting the complexities of CSRK and associated challenges for AI.

10.2 Logical complexity and the horse’s head scene in *The Godfather*

DM’s assertion that “... situations that seem straightforward can turn out, on examination, to have considerable logical complexity.” [5, p. 97] is certainly true, and the example that they give—the horse’s head scene in *The Godfather*—illustrates the point very well. In that connection, they write “Anyone who has seen [that unforgettable scene] immediately realizes what is going on. It is not just [that] it is unusual to see a severed horse head, it is clear Tom Hagen is sending Jack Woltz a message—if I can decapitate your horse, I can decapitate you; cooperate, or else. For now, such inferences lie far beyond anything in artificial intelligence.” [5, p. 93]

Since this is an interesting and challenging example, this subsection expands on how the example may be analysed and how it may be modelled with the SP system.

In summary, the relevant part of the plot is this:

“Johnny Fontane, a famous singer and godson to Vito [Corleone—the Godfather], seeks Vito’s help in securing a movie role; Vito dispatches his consigliere, Tom Hagen, to Los Angeles to talk the obnoxious studio head, Jack Woltz, into giving Johnny the part. Woltz refuses until he wakes up in bed with the severed head of his prized stallion.” (Adapted from “The Godfather”, *Wikipedia*, bit.ly/2c5YZAy, retrieved 2016-09-12.)

Instead of trying to understand the example from the perspective of a cinema audience, our analysis will focus on how Jack Woltz might interpret the unpleasant experience of finding a horse’s head in his bed. Although recognition and inference are intimately related (Section 4.2), it seems there would be those two main phases in Woltz’s thinking:

1. *Recognition.*

- (a) In order to make sense of the event, the first step is that Woltz must recognise the horse’s head as what it is. This may seem too easy and simple to deserve comment but that should not disguise the existence of this first step or its complexity.
- (b) The next step, which may again seem too simple to deserve comment, is that Woltz would make the very obvious inference that the horse’s head had been part of a horse.
- (c) Woltz would also recognise that the horse was his prized stallion which, we shall suppose, was called “Lightning Force”. We shall suppose also that a white flash on the horse’s forehead is distinctive for the stallion, although indirect inferences would probably also lead to the same identification.

2. *Inference.* Why should the head of Lightning Force have appeared in Woltz’s bed? Here are some possibilities.

- (a) It could have been some kind of accident, although it is much more likely that it was the deliberate act by some person.

- (b) Assuming that it was a deliberate act, what was the motivation? Here, Woltz’s knowledge of the Mafia would kick in: killing things is something that the Mafia do as a warning or means of persuading people to do what they want. The person to be persuaded must have an emotional attachment to the person or animal that is killed.¹⁰
- (c) Woltz also knows that Tom Hagen is a member of the Mafia and that Hagen wants Woltz to give Johnny Fontane a part in a movie. From that knowledge and his knowledge of how the Mafia operate, Woltz can make connections with the killing of Lightning Force.

Figure 5 shows how Phase 1 in the scheme above (the recognition phase) may be modelled via the creation of a multiple alignment by the SP computer model. In this example, the computer model has been supplied with a set of Old patterns describing various aspects of horses, mammals, and of Lightning Force. It has also been supplied with New information describing some of the features of horses and of Lightning Force that Woltz would have seen, and the fact that the horse was dead. Strictly speaking, Woltz would have had to infer that the horse was dead but this analysis makes the simplifying assumption that Woltz could see directly that the horse was not alive.

¹⁰This is a little different from DM’s interpretation: “... it is clear Tom Hagen is sending Jack Woltz a message—if I can decapitate your horse, I can decapitate you; cooperate, or else.” [5, p. 93] but, arguably, equally valid.

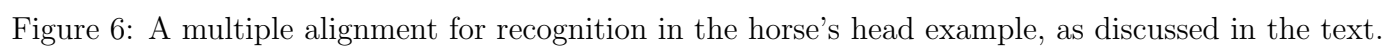
0	1	2	3	4	5
					<lf>
					lf1
					Lightning
					Force
			<horse> -----		<horse>
			h1		
		<mammal> -----	<mammal>		
		m1			
		<head> -----		<head>	
				h2	
				<marks> ----	<marks>
white-flash -----					white-flash
				</marks> ----	</marks>
long-snout -----				long-snout	
large-teeth -----				large-teeth	
		</head> -----		</head>	
		<body> -----	<body>		
			hindgut		
			fermentation		
		</body> -----	</body>		
		<legs> -----	<legs>		
			odd-toed		
		</legs> -----	</legs>		
	<vital-signs> --	<vital-signs>			
	v2				
dead -----	dead				
	</vital-signs> -	</vital-signs>			
		</mammal> -----	</mammal>		
			</horse> -----		</horse>
					</lf>
0	1	2	3	4	5

Figure 5: A multiple alignment, created by the SP computer model, for the recognition phase in the horse's head example, as discussed in the text.

In the multiple alignment, the New information, which appears in column 0, makes connections with various parts of the Old patterns in columns 1 to 5. The multiple alignment shows that the horse's head, represented by the pattern in column 4, has been recognised, that it connects with the 'head' part of a pattern representing the structure of mammals (column 2), that this pattern connects with a pattern representing horses (column 3), and that this in turn connects with a pattern representing Lightning Force (column 5). As mentioned above, we shall assume that Woltz recognises his prized stallion by the distinctive white flash on its forehead and it is this feature which brings the pattern for Lightning Force into the multiple alignment.

Figure 6 shows a multiple alignment for Phase 2 in the scheme above (the inference phase), ignoring the possibility (Phase 2 (a)) that the horse's head in Woltz's bed was the result of some kind of accident.

In principle, there could be one multiple alignment for both recognition and inference but this would have been too big to show on one page. So it has been convenient to split the analysis into two multiple alignments corresponding to the posited two phases in Woltz's thinking.



Probably the most important feature of the multiple alignment shown in Figure 6 is the pattern shown in column 3 which describes a supposed feature of how the Mafiosi operate. This is, reading from the top, that if x loves z , a member of the Mafiosi may persuade x to do something (an ‘action’ in the pattern) by killing z (the thing that x loves). This is, no doubt, a distortion and oversimplification of how the Mafiosi operate but it is perhaps good enough for present purposes.

Other features of the multiple alignment include:

- The pattern for Tom Hagen (in column 7) connects with the pattern for Mafiosi (column 3) and thus inherits their modes of operation.
- The pattern in column 13 shows that Jack Woltz (with the reference code ‘psn2’ in the pattern for Jack Woltz in column 12) ‘loves’ Lightning Force (with the reference code ‘lf1’ in the pattern for Lightning Force in column 10).
- That fact (that Jack Woltz loves Lightning Force) connects with “ x loves z ” in the pattern in column 3.
- Reading from the top, the pattern in column 8 records the fact that Tom Hagen (with the reference code ‘psn3’) is seeking to ‘persuade’ Jack Woltz (with the reference code ‘psn2’) to perform a particular ‘action’ (with the reference code ‘ac2’). That action is to “Give Johnny the part”.

The analysis of the horse’s head scene that has been presented in this section is certainly not the last word, but I believe it suggests a possible way forward. Probably the main advance that is needed for any approach, including the SP system, is robust capabilities for the unsupervised learning of CSK in realistic settings so that CSR may operate with relatively large and well-structured bodies of knowledge.¹¹

10.3 Plausible reasoning

“... commonsense reasoning almost always involves plausible reasoning; that is, coming to conclusions that are reasonable given what is known, but not guaranteed to be correct. Plausible reasoning has been extensively studied for many years, and many theories have been developed, including probabilistic reasoning, belief revision, and default reasoning or non-monotonic logic. However, overall we do not seem to be very

¹¹Preparing the example in Figure 6 has highlighted two weaknesses in the SP computer model as it is now: the need for improvements in how alternative multiple alignments are scored; and the need for more constraints in the way multiple alignments are built.

close to a comprehensive solution. Plausible reasoning takes many different forms, including using unreliable data; using rules whose conclusions are likely but not certain; default assumptions; assuming one's information is complete; reasoning from missing information; reasoning from similar cases; reasoning from typical cases; and others. How to do all these forms of reasoning [perform] acceptably well in all common-sense situations and how to integrate these different kinds of reasoning are very much unsolved problems." [5, p. 98].

In the following two subsections, I argue, first, that the SP system shows promise as a means of modelling the kinds of reasoning mentioned in the quotation, and, second, that it promises to solve the problem of integration, mentioned at the end of the quotation.

10.3.1 Modelling different kinds of reasoning

Some of the kinds of reasoning mentioned in the quotation above as being actual or potential elements of CSR seem to be the same as or similar to kinds of reasoning that have been demonstrate with the SP system (Section 5). There are reasons to believe that most of the kinds of reasoning mentioned by DM may be accommodated by the SP system:

- *Probabilistic reasoning.* As noted in Appendix A, the SP system is fundamentally probabilistic, so all kinds of reasoning in the SP system are probabilistic—although the system can, if required, imitate the clockwork nature of logical reasoning (*ibid.*).
- *Belief revision.* Since the SP system has strengths in the modelling of non-monotonic reasoning (Section 5.1), and since nonmonotonic reasoning has elements of belief revision (learning that Tweety is a penguin leads us to revise our earlier belief that Tweety can fly), there are reasons to believe that the SP system may serve to model other aspects of belief revision.
- *Default reasoning or non-monotonic logic.* As just noted, and described in Section 5.1, the SP system, with appropriate patterns, may model nonmonotonic reasoning.
- *Reasoning using unreliable data.* A general feature of the SP system is that it can deliver plausible results with New information containing errors of omission, commission, and substitution. This has been demonstrated with the parsing of natural language [15, Section 4.2.2 and Figure 6] and with pattern recognition [14, Sections 6.2 and 6.4.3]. Because of the generality

of this feature in the SP system, there is reason to believe that it will also apply in reasoning.

- *Reasoning using rules whose conclusions are likely but not certain.* Because of the probabilistic nature of the SP system (Appendix A), most of the rules used in reasoning with the system would, normally, have conclusions that are likely but not certain.
- *Reasoning with default assumptions.* The capabilities of the SP system with nonmonotonic reasoning (Section 5.1) demonstrate how it can reason with default assumptions, such as the assumption that, without contrary evidence, we may assume that if Tweety is a bird then he or she can fly.
- *Reasoning assuming one’s information is complete.* If, in relevant databases, a travel agent cannot find a direct flight between two cities, then he or she would normally tell the customer that such a flight does not exist. In a similar way, the SP system would normally be used with the “negation as failure” assumption that, if information cannot be found within the system, then that information does not exist.¹²
- *Reasoning from missing information.* This aspect of reasoning has not yet been explored in the SP programme of research. The SP theory and the SP computer model probably need to be augmented to accommodate the notion of “missing information”—because that notion provides a means of encoding information economically, in keeping with the principles on which the SP system is founded, but not yet incorporated in the SP system. To see why the concept of missing information provides a means of encoding information economically, consider how one would record the names of 10 people, all of whom are in the village football team. One could of course list them individually. But it would be more economical to record the list as something like “Bloomfield Rovers, without Jack”, where Jack is the member of the football team who is not on the list.
- *Reasoning from similar cases, and reasoning from typical cases.* Where there are similarities amongst a set of cases, or where one or more cases can be recognised as being “typical” on the strength of similarities across the range of cases, then unsupervised learning in the SP system would identify redundancies amongst the several cases and, via lossless information compression, create an abstract representation or “grammar” for those cases. That grammar would provide the basis for reasoning with those cases and, since the grammar would normally generalise beyond the cases it was derived from

¹²See “Negation as failure”, *Wikipedia*, bit.ly/2c0Ni36, retrieved 2016-09-10.

[13], it would normally provide the basis for reasoning with other cases that may be described by the grammar.

Further evidence that the SP system has potential as a vehicle for CSR may be seen in its strengths and potential in other kinds of reasoning which have the flavour of CSR, not mentioned in the quotation above but amongst those mentioned in Section 5:

- *Bayesian reasoning with “explaining away”.*
- *Causal reasoning.*
- *Reasoning that is not supported by evidence.*
- *Inference via inheritance of attributes.*
- *Spatial reasoning.*
- *What-if reasoning.*

10.3.2 Seamless integration of CSR

As described in Section 6 and elsewhere in this paper, the use of one very simple format for the representation of all kinds of knowledge and one relatively simple framework—multiple alignment—for the processing of knowledge, are likely to facilitate the seamless integration of diverse kinds of knowledge and diverse aspects of intelligence.

These remarks apply with equal force to the several forms of reasoning within the actual or potential capabilities of the SP system (Sections 5 and 10.3.1).

As noted in Section 6, ordinary experience suggests that seamless integration of diverse kinds of knowledge and diverse forms of reasoning are pre-requisites for the kinds of commonsense reasoning which we do constantly in many different situations.

10.4 Long tail

“... in many domains, a small number of examples are highly frequent, while there is a ‘long tail’ of a vast number of highly infrequent examples.” (*ibid.*).

Probably the most famous example of the “long tail” phenomenon is the sentence *Colorless green ideas sleep furiously* that Noam Chomsky [4, p. 15] presented to illustrate, *inter alia*, how a sentence that is grammatical can be vanishingly rare.

It is widely accepted that most sentence in most natural languages are new to the world.¹³

On the strength of examples like “Colorless green ideas ...”, and the extraordinary complexity of natural language, Chomsky and others developed the ‘nativist’ view that much of the structure of natural language is inborn.¹⁴ However, models of the unsupervised learning of language via the compression of information demonstrate how it is possible to develop a knowledge of structural elements like words, classes of words, and abstract patterns [13]. Chomsky’s assertion that “... one’s ability to produce and recognize grammatical utterances is not based on notions of statistical approximation and the like.” [4, p. 16] is, almost certainly, too strong.

An important point for the present discussion is that unsupervised learning via information compression, which includes unsupervised learning in the SP system, provides a persuasive account of how it is possible, without correction by a ‘teacher’ or equivalent assistance, to create a grammar that generalises beyond the raw data without over-generalising ([13, pp. 181–191], [14, Chapter 9], [15, Section 5.3]). In brief, the products of such learning are: a *grammar* which provides an abstract description of the raw data; and an *encoding* of the raw data in terms of the grammar. The two things together provide lossless compression of the raw data. And, normally, the grammar generalises beyond the raw data without overgeneralising.

The significance of these observations for the long-tail phenomenon and CSR, which applies to any kind of data, not just natural language, is that generalisations beyond the raw data, of which there may be many, are likely to be very rare in most samples of raw data, or entirely absent from all but the very largest samples of such data.

But it appears that the rarity of many generalisations is of little or no consequence for CSR, provided that the grammar captures recurrent features of the world, as would normally be the case with unsupervised learning via information compression. Consider, for example, the commonsense argument that if **A** comes before **B**, and **B** comes before **C**, then **A** comes before **C**. This kind of relationship is a recurrent feature of the world, regardless of the rarity of **A**, **B**, and **C**, individually or in combination. Those things might, for example, be the very rare combination **elephant**, **screwdriver**, and **veggie burger** but the rarity of such a combination

¹³In brief, this can be proved as follows. Since recursive structures are prominent in most natural languages, and since such structures can produce infinitely many surface forms, there is an infinite number of different possible sentences in any natural language. This is larger, by a wide margin, than the finite albeit large number of sentences that have actually been written or spoken.

¹⁴See, for example, “Psychological nativism”, *Wikipedia*, bit.ly/1ePaAp4, retrieved 2016-09-12.

does not disturb the general rule of which it is an example.

10.5 Discerning the proper level of abstraction

“... in formulating knowledge it is often difficult to discern the proper level of abstraction. Recall the example of sticking a pin into a carrot and the task of reasoning that this action may well create a hole in the carrot but not create a hole in the pin. ... The question is, how broadly should such rules should be formulated?” [5, pp. 98–99].

In brief, the putative “SP” answer to the “proper level of abstraction” and “how broadly should ... rules ... be formulated” is compression information. In keeping with the argument in Section 9.4—that we should not attempt to determine knowledge structures via analysis but should be guided by what emerges from a well-constructed learning system founded on ICMA—we should see what levels of abstraction emerge from learning via ICMA. By hypothesis, these would represent the proper levels of abstraction, where criteria for “proper” would include succinctness and naturalness in CSK (in accordance with the DONSVIC principle) and effectiveness and efficiency in CSR.

10.6 Methodological and sociological obstacles

“A final reason for the slow progress in automating commonsense knowledge is both methodological¹⁰ and sociological. Piecemeal commonsense knowledge (for example, specific facts) is relatively easy to acquire, but often of little use, because of the long-tail phenomenon discussed previously. Consequently, there may not be much value in being able to do a little commonsense reasoning.” [5, p. 99].

The main points in this subsection are puzzling. It’s not clear why piecemeal commonsense knowledge should be of little use or why the long-tail phenomenon is relevant. A young child is likely to learn quickly that getting burned is painful but food is normally nice, and such knowledge is likely to prove useful despite the fact that they are specific facts. And they connect with frequently-occurring situations that are not in the “long-tail” category.

11 Objectives for research in CSRK

This section contains some brief comments on how the SP system relates to DM’s “objectives for research in commonsense reasoning” [5, pp. 99–100]:

- *Reasoning architecture.* This means “The development of general-purpose data structures for encoding knowledge and algorithms and techniques for carrying out reasoning.”. In the SP system, the multiple alignment framework with SP patterns has proved to be a versatile system for the representation of knowledge (Section 3) and for reasoning (Section 5).
- *Plausible inference.* “Drawing provisional or uncertain conclusions” is central in the workings of the SP system since the system is fundamentally probabilistic (Appendix A).
- *Range of reasoning modes.* With regard to this objective—“incorporating a variety of different modes of inference, such as explanation, generalization, abstraction, analogy, and simulation”:
 - “Explanation” is an implicit part of unsupervised learning in the SP system since a main product of that learning is a “grammar” which may be regarded as theory of raw data from which interpretations or explanations may be drawn via the building of multiple alignments.
 - “Generalisation” is an important part of unsupervised learning by the SP system as outlined in Section 10.4.
 - “Abstraction” is a fundamental part of unsupervised learning by the SP system.
 - “Analogy” has not been addressed directly in the SP programme of research, but the SP system is clearly relevant to this topic because of its ability to recognise similarities between patterns, outlined in Section 4.1.
 - Again, “simulation” has not been an explicit focus of interest in the SP programme to date, but the system is relevant to that topic because a grammar that has been abstracted from a given body of raw data via unsupervised learning provides a means of simulating the source or sources of those data.
- *Painstaking analysis of fundamental domains.* “In doing commonsense reasoning, people are able to do complex reasoning about basic domains such as time, space, naïve physics, and naïve psychology. The knowledge they are drawing on is largely un verbalized and the reasoning processes largely unavailable to introspection. An automated reasoner will have to have comparable abilities.” Clearly, careful analysis of human CSRK will be needed for the successful automation of such knowledge representation and reasoning in artificial systems, including the SP system.

- *Breadth.* “Attaining powerful commonsense reasoning will require a large body of knowledge.” This is clearly true for artificial systems including the SP system.
- *Independence of experts.* “Paying experts to hand-code a large knowledge base is slow and expensive. Assembling the knowledge base either automatically or by drawing on the knowledge of non-experts is much more efficient.” The strengths and potential of the SP system in unsupervised learning is likely to prove useful in the automatic learning of knowledge. Learning from books and other written material is likely to be important for CSRK and here the strength and potential of the SP system in the interpretation of natural language is clearly relevant, although substantial work will be needed to develop true understanding of text, not the relatively superficial processing in IBM’s Watson, mentioned by DM [5, p. 94] (see also [20, Section IX]).
- *Applications.* “To be useful, the commonsense reasoner must serve the needs of applications and must interface with them smoothly.” Since it is envisaged that, in mature versions of the SP system, all applications and CSRK will be realised via SP patterns in the multiple alignment framework, and since the one simple format for knowledge and the one relatively simple framework for the processing of knowledge is likely to facilitate the seamless integration of knowledge and processing (Section 6), there are reasons to believe that the SP system will facilitate the smooth interfacing of the commonsense reasoner with diverse applications, all of them hosted on the SP system.
- *Cognitive modelling.* The SP programme of research is founded on earlier research that highlights the significance of information compression in the workings of brains and nervous systems and in children’s learning of natural language (Appendix A).

12 Conclusion

Understanding commonsense reasoning and commonsense knowledge is indeed challenging, but the *SP theory of intelligence* and its realisation in the *SP computer model* have relevant strengths and potential. In brief:

- *The generality of a universal Turing machine.* It appears that the SP system has the generality of a universal Turing machine, the kind of generality that is a pre-requisite for CSRK.

- *Generality in information compression via multiple alignment.* Likewise, information compression via multiple alignment, which is central in the workings of the SP system:
 - Provides the generality needed for *the representation of diverse forms of knowledge*, and, via the DONSVIC principle, it provides for *succinctness* in those representations.
 - And, owing to the intimate relation between information compression and concepts of prediction and probability, *inference and probabilities are embedded in the workings of the SP system*, in keeping with the probabilistic nature of CSR.
- *Versatility and integration.* Despite the relative simplicity of the SP system with multiple alignment centre stage, the system has substantial versatility in areas that are needed for CSRK:
 - *Versatility in the representation of knowledge.* SP patterns, within the multiple alignment framework, have proved to be effective in representing several forms of knowledge, any of which may serve in CSK: the syntax of natural language; class hierarchies, class heterarchies (meaning class hierarchies with cross classification); part-whole hierarchies; discrimination networks and trees; entity-relationship structures; relational knowledge; rules for use in reasoning; patterns in one or two dimensions; images; structures in three dimensions; and procedural knowledge.
 - *Versatility in aspects of intelligence.* The SP system has strengths and potential in several aspects of intelligence: unsupervised learning; natural language processing; fuzzy pattern recognition; recognition at multiple levels of abstraction; best-match and semantic forms of information retrieval; several kinds of reasoning (more next); planning; and problem solving.
 - *Versatility in reasoning.* Strengths and potential of the SP system include: one-step ‘deductive’ reasoning; chains of reasoning; abductive reasoning; reasoning with probabilistic networks and trees; reasoning with ‘rules’; nonmonotonic reasoning and reasoning with default values; Bayesian reasoning with “explaining away”; causal reasoning; reasoning that is not supported by evidence; inheritance of attributes; spatial reasoning; and what-if reasoning.
 - *Seamless integration of diverse forms of knowledge and diverse aspects of intelligence.* The use of one simple format for knowledge and one

relatively simple framework for the processing of knowledge promotes seamless integration of diverse forms of knowledge and diverse aspects of intelligence, an integration that appears to be essential for CSRK.

- *Examples of CSR.* The paper discusses several of DM’s examples of CSR, with multiple alignments for recognition and inferences that may arise in the horse’s head scene in *The Godfather*.
- *Successes in automated commonsense reasoning.* Also discussed are current successes in CSR (taxonomic reasoning, temporal reasoning, action and change, and qualitative reasoning), how the SP system may promote seamless integration across these areas, and how insights gained from the SP programme of research may yield some potentially useful new ways of approaching these topics. Insights gained in those areas may prove useful in future development of the SP system.
- *Challenges in automating commonsense reasoning.* The paper considers how the SP system may help overcome some of the challenges, described by DM, in the automation of CSR:
 - The logical complexity of much of CSR (including the afore-mentioned horse’s head example).
 - The SP system shows promise as a means of modelling the kinds of “plausible reasoning” mentioned by DM, and several others that they don’t mention. And the SP system promises to solve the problem of integration in plausible reasoning that DM mention.
 - Research on the unsupervised learning of natural language, which has provided much of the inspiration for the SP programme of research, provides insights into the “long tail” phenomenon—the existence, in most domains, of many examples that occur only very infrequently—and, in particular, why the existence of such rare examples would normally be of little or no consequence for CSR.
 - It is envisaged that, when the SP system is more mature, levels of abstraction for reasoning and other aspects of intelligence would be determined via unsupervised learning in accordance with the DONSVIC principle—the discovery of natural structures via information compression;
- *Objectives for research in CSRK.* The SP system has what appear to be useful things to say about several of DM’s objectives for research in CSRK: the

development of a general-purpose reasoning architecture; how to draw provisional or uncertain conclusions; how to incorporate a variety of different modes of inference; how reasoning may integrate smoothly with applications; and the need for consistency with human cognition. In agreement with what DM say, success with CSRK will require painstaking analysis of different areas of CSRK; large bodies of knowledge will be needed for success in modelling CSRK; and it would be too slow and expensive to glean relevant knowledge from experts.

A Outline of the SP system

This is a bare-bones outline of the SP system. More information, with increasing levels of detail, may be found in [20, Appendix I], [15, Sections 3, 4 and 5], and [14, Chapters 3 and 9].

The *SP theory of intelligence* and its realisation in the *SP computer model* is the product of about 20 years of research, seeking to discover or invent a conceptual framework that simplifies and integrates observations and concepts across artificial intelligence, mainstream computing, mathematics, and human perception and cognition.

Distinctive features of the SP system and its advantages compared with several AI-related alternatives are described in [20]. In particular, Section V describes several problems with deep learning in artificial neural networks—the subject of much interest at present—and how, in the SP framework, those problems may be overcome.

Other key papers in the SP programme of research, including several about potential benefits and applications of the system, are detailed with download links near the top of www.cognitionresearch.org/sp.htm.

Key features of the SP system are:

- *Founded on research in neuroscience and cognitive science.* The SP programme of research has its origins in research on the role of information compression in the workings of brains and nervous systems by Attneave [1], Barlow [2, 3] and others, and in my own research developing computer models of language learning by children (summarised in [13]), in which information compression proved to be of central importance.
- *The representation of knowledge with SP ‘patterns’.* By hypothesis, all kinds of knowledge may be represented in the SP system with arrays of atomic symbols in one or two dimensions called *patterns*. At present, the SP computer model works only with one-dimensional patterns but it is envisaged

that the model will be generalised to work with patterns in two dimensions [15, Section 3.3].

- *Processing via the matching and unification of patterns.* By hypothesis, all kinds of processing in the SP system may be done via “information compression via the matching and unification of patterns” (ICMUP), where “patterns” includes parts of patterns as well as whole patterns.
- *Processing via multiple alignments.* More specifically, it is envisaged that all kinds of processing in the SP system is done via “information compression via the building and processing of multiple alignments” (ICMA), where *multiple alignment* is a concept borrowed and adapted from bioinformatics. As noted in the Introduction, multiple alignment, as it has been developed in the SP programme of research, has the potential to be as significant for an understanding of intelligence in a broad sense as is DNA for biological sciences.
- *Unsupervised learning.* Unsupervised learning in the SP system is achieved by direct assimilation of “New” information from the system’s environment (to create “Old” patterns for storage by the system), by creating Old patterns from multiple alignments in which there are partial matches between patterns, and via heuristic search through alternative *grammars* (collections of Old patterns), to find one or two that score better than others in terms of information compression. At present, the SP computer model demonstrates unsupervised learning of plausible generative grammars for the syntax of English-like artificial languages, including the learning of segmental structures, classes of structure, and abstract patterns, but further work is needed to realise the system’s full potential in this area [15, Section 3.3].
- *Other aspects of intelligence.* Aspects of intelligence other than unsupervised learning, such as pattern recognition, information retrieval, several kinds of reasoning, and more, are modelled in the SP system via the building of multiple alignments. Two examples showing how the parsing of natural language may be modelled via the building of multiple alignments is shown below. Other examples of multiple alignments are shown in the main part of the paper.
- *The DONSVIC principle.* A key idea in the SP framework is that the entities and abstract concepts discovered via unsupervised learning via multiple alignment would be forms of knowledge that people recognise as “natural”, including specific entities like “my cat” and classes of such entities like “animal”. Evidence to date shows that the SP computer model conforms to this

principle—the discovery of natural structures via information compression, or “DONSVIC” for short [15, Section 5.2]. Empirical evidence from the SP computer model, and analysis of how it works, suggests that the discovery of natural structures goes hand-in-hand with the achievement of relatively high levels of information compression.

- *The SP system is fundamentally probabilistic.* Because of the intimate connection that is known to exist between information compression and concepts of prediction and probability [6], the whole system is fundamentally probabilistic. Each SP pattern has an associated frequency of occurrence which provides the basis for the calculation of an absolute and relative probability for each multiple alignment and for every inference that may be drawn from any multiple alignment.

Although the SP system is fundamentally probabilistic, it can, via the processing of forms of knowledge that yield probabilities equal to or close to the values 0 and 1, imitate the all-or-nothing clockwork nature of traditional computing ([15, Section 6.3], [14, Chapter 10]).

Figure 7 shows two examples of multiple alignment, demonstrating alternative syntactic parsings of the ambiguous sentence *fruit flies like a banana*.¹⁵

In each multiple alignment, the sentence to be parsed is shown as a “New” pattern, meaning that it is input to the system. By convention New information is always shown in row 0.

In each multiple alignment, grammatical structures, including words, are represented as SP patterns in rows 1 to 8, one pattern per row. The order of the patterns across the rows has no significance. These are a few of a relatively large set of “Old” patterns, meaning patterns that have been stored in the system prior to the input and analysis of the New pattern. In these two examples, the set of Old patterns is, in effect, a grammar for the syntactic analysis of natural language sentences.

The SP computer model builds multiple alignments like these by pairwise alignment of SP patterns and previously-formed alignments, in much the same way as programs for the building of multiple alignments in bioinformatics. In both cases, the abstract space of possible multiple alignments is far too big to be searched exhaustively so it is necessary to use heuristic techniques, searching for “good” multiple alignments in stages and pruning the search tree at each stage.

The main difference between multiple alignments in the SP system and those in bioinformatics is that, in the latter, all rows have the same status while in SP multiple alignments, row 0 contains a New pattern (sometimes more than one),

¹⁵This sentence is the second part of *Time flies like an arrow. Fruit flies like a banana.*, attributed to Groucho Marks.

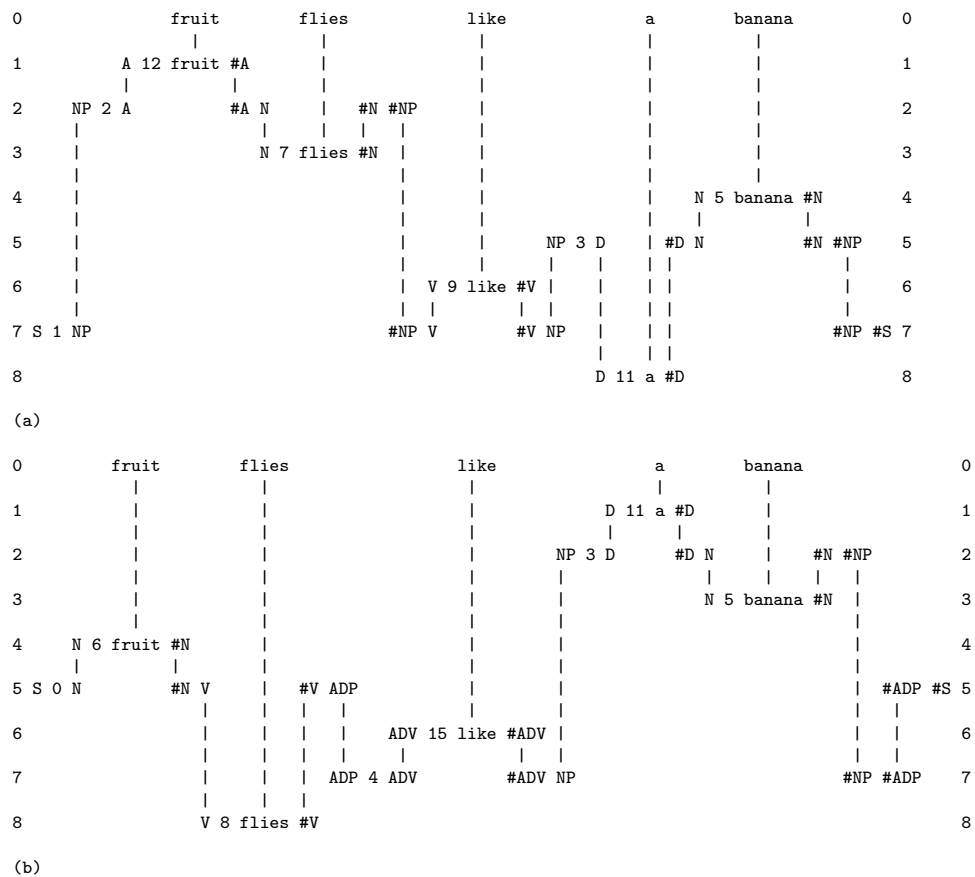


Figure 7: The two best multiple alignments created by the SP computer model showing two different parsings of the ambiguous sentence *Fruit flies like a banana* in terms of SP patterns representing grammatical categories, including words. Here, multiple alignments are evaluated in terms of economical encoding of information as outlined in the text. Adapted from Figure 5.1 in [14], with permission.

other rows contain Old patterns, one per row, and the system aims to find one or more multiple alignments that enable to New pattern to be encoded economically in terms of the Old patterns, as described in [14, Section 3.5] and [15, Section 4.1].

These examples merely demonstrate how the SP system may achieve syntactic parsing of natural language. Preliminary examples showing how meanings may be derived from the surface forms of language and how surface forms may be derived from meanings are shown in [14, Section 5.7].

The two multiple alignments in the figure show how the SP system is able to find alternative interpretations of a given body of information.

References

- [1] F. Attneave. Some informational aspects of visual perception. *Psychological Review*, 61:183–193, 1954.
- [2] H. B. Barlow. Sensory mechanisms, the reduction of redundancy, and intelligence. In HMSO, editor, *The Mechanisation of Thought Processes*, pages 535–559. Her Majesty’s Stationery Office, London, 1959.
- [3] H. B. Barlow. Trigger features, adaptation and economy of impulses. In K. N. Leibovic, editor, *Information Processes in the Nervous System*, pages 209–230. Springer, New York, 1969.
- [4] N. Chomsky. *Syntactic Structures*. Mouton, The Hague, 1957.
- [5] E. Davis and G. Marcus. Commonsense reasoning and commonsense knowledge in artificial intelligence. *Communications of the ACM*, 58(9):92–103, 2015.
- [6] M. Li and P. Vitányi. *An Introduction to Kolmogorov Complexity and Its Applications*. Springer, New York, 3rd edition, 2014.
- [7] D. Marr. *Vision: a Computational Investigation into The Human Representation and Processing of Visual Information*. The MIT Press, London, England, 2010. This book was originally published in 1982 by W. H. Freeman and Company.
- [8] P. McCorduck. *Machines who think: a personal inquiry into the history and prospects of artificial intelligence*. A. K. Peters Ltd, Natick, MA, second edition, 2004. ISBN: 1-56881-205-1.

- [9] J. Pearl. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. Morgan Kaufmann, San Francisco, revised second printing edition, 1997.
- [10] A. M. Turing. Computing machinery and intelligence. *Mind*, 59:433–460, 1950.
- [11] C. S. Webster. Alan turing’s unorganized machines and artificial neural networks: his remarkable early work and future possibilities. *Evolutionary Intelligence*, 5:35–43, 2012.
- [12] T. Winograd. Understanding natural language. *Cognitive Psychology*, 3(1):1–191, 1972.
- [13] J. G. Wolff. Learning syntax and meanings through optimization and distributional analysis. In Y. Levy, I. M. Schlesinger, and M. D. S. Braine, editors, *Categories and Processes in Language Acquisition*, pages 179–215. Lawrence Erlbaum, Hillsdale, NJ, 1988. bit.ly/ZIGjyc.
- [14] J. G. Wolff. *Unifying Computing and Cognition: the SP Theory and Its Applications*. CognitionResearch.org, Menai Bridge, 2006. ISBNs: 0-9550726-0-3 (ebook edition), 0-9550726-1-1 (print edition). Distributors, including Amazon.com, are detailed on bit.ly/WmB1rs.
- [15] J. G. Wolff. The SP theory of intelligence: an overview. *Information*, 4(3):283–341, 2013. bit.ly/1hz0lFE.
- [16] J. G. Wolff. Application of the SP theory of intelligence to the understanding of natural vision and the development of computer vision. *SpringerPlus*, 3(1):552–570, 2014. bit.ly/1scmpV9.
- [17] J. G. Wolff. Autonomous robots and the SP theory of intelligence. *IEEE Access*, 2:1629–1651, 2014. bit.ly/1zrSemu.
- [18] J. G. Wolff. Big data and the SP theory of intelligence. *IEEE Access*, 2:301–315, 2014. bit.ly/1jGWXDH. This article, with minor revisions, is to be reproduced in Fei Hu (Ed.), *Big Data: Storage, Sharing, and Security (3S)*, Taylor & Francis LLC, CRC Press, 2016, pp. 143–170.
- [19] J. G. Wolff. The SP theory of intelligence: benefits and applications. *Information*, 5(1):1–27, 2014. bit.ly/1lcquWF.
- [20] J. G. Wolff. The SP theory of intelligence: its distinctive features and advantages. *IEEE Access*, 4:216–246, 2016. bit.ly/21gv2jT.